

● 王 宁¹, 赵胜洋¹, 单晓红²

(1. 北京邮电大学经济管理学院, 北京 100876; 2. 北京工业大学经济与管理学院, 北京 100022)

基于灰色系统理论的网络舆情预测与分级方法研究*

摘要: [目的/意义] 建立一套行之有效的网络舆情预测和分级管理系统对提高政府网络监管效率具有重大意义。[方法/过程] 分时段记录网络舆情事件的新浪微指数、百度指数、头条指数作为事件热度的衡量指标, 运用 EGM(1.1) 模型对舆情事件的发展趋势进行预测, 并在预测数据基础上, 运用灰色关联分析方法, 提出网络舆情事件分级方案。综合运用上述两模型确立的舆情监管体系对“江歌案宣判”“上海携程亲子园虐童案”“莫焕晶案二审宣判”等事件进行分析评价。[结果/结论] 模型在处理舆情管理问题方面具有良好的适用性和精确性。结合舆情事件特点与分析结果, 向政府及公司危机公关部门等提出了相应建议。

关键词: 网络舆情; 灰色系统理论; 舆情预测和分级; 实证研究

Method of Network Public Opinion Prediction and Grading Based on Grey System Theory

Abstract: [Purpose/significance] Establishing an effective system to predict and grade network public opinion has great significance in improving the efficiency of governmental network supervision. [Method/process] First, record Sina micro index, Baidu index, Toutiao index of network public opinion events hour by hour which will be used as the measurement indices of event concern degree. Then EGM(1.1) Model is used to predict the development trend of public opinion events and grey relational analysis is applied to propose a grading scheme for network public opinion events based on the predicted data. The paper analyzes and evaluates typical events such as the “Jiang Ge case” “Shanghai Xiecheng child abusing case” “Mo Huanjing case” with the public opinion supervision system established by the two models above. [Result/conclusion] Established models have great applicability and precision in dealing with the issues of public opinion management. Combined with the characteristics of public opinion events and the system analysis results, suggestions oriented the government and the company’s public relation departments have been made in the paper.

Keywords: network public opinion; grey system theory; network public opinion prediction and grading; empirical study

随着时代的发展, 互联网越来越成为表达民意、反映舆情的重要平台, 尤其是“微博”“微信”等“自媒体”社交软件的快速发展, 使得网络舆情成为社会舆情环境的重要组成部分。与传统舆情相比, 网络舆情具有不确定性更大、传播速度更快、反应时间更短等特征^[1], 这意味着在网络舆情事件爆发前后, 能利用的数据仅包括几小时内的少量数据, 可利用信息非常贫乏, 事件发展趋势, 群众态度倾向等更重要的舆情方面也存在不确定性大的问题。这不仅给广大研究学者抛出了一个难题, 也在很大程度上增加了政府等管理部门对于舆情事件的监管和把控难度。

基于此, 本文提出运用灰色系统理论知识, 构建基于

EGM(1.1) 模型和灰色关联分析方法的网络舆情预测和分级管理系统。灰色系统理论研究对象是“部分信息已知, 部分信息未知”的“小样本”“少数据”“贫信息”不确定系统^[2]。结合前文提到的网络舆情“可利用信息贫乏”“不确定性大”等特点, 可以得出: 舆情事件监管相关系统符合灰色系统特征, 适用灰色系统理论相关算法与模型。灰色系统理论能够提供解决网络舆情事件“数据少”“不确定性高”等新思路。

1 文献回顾

目前, 关于如何解决网络舆情事件的预测和分级管理问题, 国内外学者做出了很多深入的研究, 提出了诸多不同的方法和模型。从现有的研究成果来看, 研究方法主要分为以下三类。

1.1 基于运筹管理思想的解决方法

基于运筹管理思想的方法主张构建舆情事件评价指标体系, 在评价指标体系的基础上, 运用德尔菲法、模糊综

* 本文为北京市自然科学基金资助项目“多重共现耦合的科技知识网络关联发现研究: 链路预测的视角”(项目编号: 9174029)和2017北京工业大学人文社会科学项目“北京突发事件网络舆情演化机理与应对策略”(项目编号: 011000546318521)的研究成果。

合评判、AHP 层次分析法等运筹管理方法,赋予不同指标不同的权重,最终得出对不同事件的热度评价。此类方法注重对事件的事后评价,根据不同的热度对事件进行分级,为管理者提供管理建议。其中,李文杰等采用分层抽样法,以网络、专家和纸质问卷调查相结合的方式得到各级指标量权重,并在此基础上通过多级模糊综合评判方法建立了舆情信息安全的模糊评判模型,实现了对舆情事件的分级与评价^[3]。刘健等根据信息传播效果的“认知—情感—行为”阶梯模式理论,利用模糊数据包络分析法建立了微博舆情信息传播效果评价指标体系,从控制传播效果的角度提出了监管建议^[4]。王高飞等将层次分析法与模糊综合分析法相结合,构建了移动社交网络舆情预警模型,并以“8·12 天津港爆炸事故”进行实证分析,证明了该模型的有效性和准确性^[5]。刘毅利用基于三角模糊数的模糊德尔菲法和模糊层次分析法,实现了对某一具体公共事务或热点话题的分析评价^[6]。

1.2 基于计算机算法的解决方法

基于计算机算法的方法将很多先进的计算机算法迁移到舆情监管的领域中,通过赋予算法参数舆情管理相关方面的现实意义,适当修改模型,实现对舆情事件的分析评价过程。此类方法成功运用了计算机反应速度快、并行处理和分布式处理能力等优点,能够在一定程度上适应舆情事件爆发速度快、反应时间短的特性。其中,梁晓贺等基于超网络分析法构建了包含用户—观点—情感—时序阶段 4 层子网的超网络模型,以达到预测舆情主题演化的目的^[7]。田世海等运用 SVM 算法完成对突发性安全事件的舆情预警工作,并应用 MATLAB 进行仿真实验,证明了模型的切实可行性^[8]。郭韧等利用可拓聚类理论对网络舆情的演化趋势进行建模并预测,通过聚类分析得到预测结果^[9]。张一文等结合贝叶斯算法的特点,建立网络舆情态势评估模型,从而对网络舆情态势进行有效评估^[10]。胡悦等以微博为研究对象,运用模糊神经网络算法,提高了微博舆情发展趋势预测的准确性^[11]。何建民等运用隐马尔科夫模型理论,选择网民特征、信息主题和信息内容完整度三维指标,实现了对微博舆情演化的预测^[12]。Yang 等提出了一套改进的 SVM 算法,并以此为基础实现了对网络事件参与者的观点预测^[13]。

1.3 基于在线数据抓取和内容分析技术的解决方法

基于在线数据抓取和内容分析技术的方法主要通过抓取监控平台内的评论信息,对采集到的文本使用文本分析方法,以达到及时发现和预测舆论发展方向,了解大众态度倾向的目的。此方法具有实时性高,能够反映大众态度倾向的优点。在此类研究中,安璐等学者采用基于词典的情感分析方法,在识别与预测网络舆情潜在问题方面提出

了自己的观点^[14]。赵晓航以新浪微博“天津爆炸”话题为例,为突发事件下政府如何利用微博平台收集、研判、应对网络舆情事件提供了科学的分析方法和应对手段^[15]。董坚峰利用 Web 挖掘技术,构建了突发事件网络舆情预警系统模型,实现了对突发事件的自动分级处理^[16]。李彤等通多对微博内容的情感分析与集成,达到了对于微博舆情趋势预测的目标^[17]。王英等利用爬虫工具和文本分析技术对微博内容进行处理,最终完成对网络舆情事件的研判过程^[18]。V. Gopalakrishnan 等通过抓取病人对药品的评论信息,运用情感分析的方法,实现了对病人意见观点的预测和分析^[19]。W. Cherif 等则是对一个旅游网站上的 625 条阿拉伯语评论进行分析评价从而达到了对使用阿拉伯语的人群进行意见分类和观点预测目标^[20]。

综上所述,目前国内外学者对如何解决网络舆情预测和分级管理问题的相关研究众多,但现有成果缺少对舆情事件预测和分级管理两个方面的综合性深入研究。部分学者注重舆情的预测环节,着重于构建能够研究舆情事件未来走势、话题演化预测、观点预测等方面的系统。这类系统往往建设复杂性高,成本预算大。部分学者注重对舆情事件的分析评价,为以后管理者处理类似舆情事件提供建议,而此类方法注重事后分析,实时性太差,无法适应网络舆情爆发快、反应时间短的特点。基于以上分析,本文主张以灰色系统理论的 EGM(1.1) 模型与灰色关联分析方法为基础,利用 EGM(1.1) 模型良好的预测性及灰色关联分析法优秀的分类准确性,构建网络舆情预测与分级管理体系,实现预测和分级的统一。此外,本文以新浪微指数、百度指数、头条指数三项指标数据构建舆情事件的衡量指标体系,三项数据采集方便且数据实时性高,一方面极大减少系统构建成本,另一方面也能很好适应舆情事件爆发快,反应时间短的特性。

2 网络舆情预测与分级管理体系构建

2.1 网络舆情衡量指标体系

利用相关指标量建立对事件热度或影响度的衡量指标体系是解决舆情事件相关监管问题的首要步骤。本文采用新浪微博的微指数、百度搜索的百度指数、今日头条的头条指数三种指数作为事件热度的评价指标。三种指数分别为新浪、百度、今日头条三家公司基于大数据处理技术建立起的大数据平台,其原理是在公司掌握的相关关键词的用户搜索数、互动数等数据基础上,运用大数据算法计算而来。指数功能推出至今,在处理舆情问题上成果不甚显著,其中陈涛等以百度指数和谷歌趋势数据为基础,实现了对“小悦悦事件”等热点时事的分析^[21]。王康等以百度指数作为公司舆情热度指标,为公司应对舆情危机提供

方法依据^[22]。张玉强构建了基于新浪微指数的阶段性事件演化分析体系,为事件的识别与演化分析提供了新思路^[23]。三大指数功能除了应用前景广阔外,还具备数据方便易得的特点,只需登录相应官方网站即可搜索获得,数据的时效性和完整性也较高。因此,本文认为以舆情事件的三大指数值为基础建立的事件热度衡量体系具有合理性和优越性。

2.2 基于 EGM(1.1) 模型的舆情预测过程

在灰色系统理论中,GM 系列模型是灰色预测理论的基本模型,尤其是邓聚龙教授提出的 GM(1.1) 模型,已在很多领域成功应用。该模型是一种时间序列预测模型,它能根据少量信息进行建模与预测,因而得到广泛的应用。GM(1.1) 模型发展至今已有很多优化和改进,其中刘思峰教授针对不同模型的性质和原始数列的特点,提出了 4 种 GM(1.1) 模型的基本形式,即:均值 GM(1.1)

模型 (EGM)、原始差分 GM(1.1) 模型 (ODGM)、均值差分 GM(1.1) 模型 (EDGM)、离散 GM(1.1) 模型 (DGM)。具体应用中应根据不同的应用情景和原始数据的特点选择不同的模型^[24]。

在舆情事件发生后,登录相关指数官方网站即可记录事件具体指数值。本文以“江歌案宣判”“上海携程亲子园虐童案”(以下简称“携程虐童案”)“莫焕晶案二审宣判”三个舆情事件为例进行模型的建立和检验过程。采集到的相关舆情事件原始数据见表 1。从表 1 中可以看出,数据序列单调性规律不强,增长区间的增长速度也没有达到指数级增长速率,基本符合振荡序列、非指数增长序列的特点。而且诸多学者也表示舆情领域的数据序列属于振荡序列^[25]。因此,根据刘思峰教授的主张:对于非指数增长序列和振荡序列,应首选微分、差分混合形态的均值 GM(1.1) 模型 (EGM)。

表 1 舆情事件原始指数数据

江歌案宣判	12/20 14 时	12/20 15 时	12/20 16 时	12/20 17 时	12/20 18 时	12/20 19 时	12/20 20 时	12/20 21 时
新浪微指数	12237	15646	18039	18153	17732	18458	19041	19367
百度指数	16569	20280	22032	20467	20142	21069	21982	22432
头条指数	421.3k	622.7k	651.4k	678.8k	623.9k	619.7k	622.1k	641.1k
携程虐童案	11/8 9 时	11/8 10 时	11/8 11 时	11/8 12 时	11/8 13 时	11/8 14 时	11/8 15 时	11/8 16 时
新浪微指数	6254	9631	10872	12078	12935	13391	14937	15406
百度指数	3935	5163	5471	5824	5964	6139	6319	6543
头条指数	331.6k	436.2k	457.8k	528.7k	535.1k	546.7k	560.4k	571.9k
莫焕晶案宣判	2/9 10 时	2/9 11 时	2/9 12 时	2/9 13 时	2/9 14 时	2/9 15 时	2/9 16 时	2/9 17 时
新浪微指数	10359	12937	13038	11022	9064	8264	7258	6841
百度指数	5722	8124	8083	7465	6953	5596	5107	4668
头条指数	472.3k	546.4k	526.7k	503.8k	482.3k	467.1k	455.3k	439.8k

本文以江歌案最终宣判事件的新浪微指数为例介绍模型构建过程。

1) 根据表 1 的江歌案最终宣判的新浪微指数数据可得原始序列为:

$$X^{(0)} = (12237, 15646, 18039, 18153, 17732, 18458, 19041, 19367)$$

2) 对原始数据进行一次累加可得:

$$X^{(1)} = (12237, 27883, 45922, 64075, 81807, 100265, 119306, 138673)$$

3) 构造数据系列 B 和数据向量 Y:

$$B = \begin{bmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \dots & 1 \\ -\frac{1}{2}(x^{(1)}(n-1) + x^{(1)}(n)) & 1 \end{bmatrix} = \begin{bmatrix} -20060 & 1 \\ -36902.5 & 1 \\ -54998.5 & 1 \\ -72941 & 1 \\ -91036 & 1 \\ -109785.5 & 1 \\ -128989.5 & 1 \end{bmatrix},$$

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{bmatrix} = \begin{bmatrix} 15646 \\ 18039 \\ 18153 \\ 17732 \\ 18458 \\ 19041 \\ 19367 \end{bmatrix}$$

4) 确定参数 a 和 b。采用最小二乘法对待定系数求解则有:

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y = \begin{bmatrix} -0.026332 \\ 16126.085341 \end{bmatrix}$$

5) EGM(1, 1) 模型的时间响应序列为:

$$x^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{u}{a} \right] e^{-ak} + \frac{u}{a} = 624651.90432e^{0.026332k} - 612414.90432$$

6) 将不同 k 值带入响应序列可得模拟值的一次累加序列,对此序列做一次累减计算,即可得模拟值序列记为 $\hat{x}^{(0)}(k)$ (为方便对比,此处取整数)(见表 2)。

表2 江歌案新浪微博指数模拟值与实际值对比

实际值	12237	15646	18039	18153	17732	18458	19041	19367
一次累加模拟值	12237	28904	46015	63583	81620	100138	119150	138670
模拟值	12237	16667	17111	17568	18037	18518	19012	19519

7) 精度检验。计算均方差比 C ：

$$\varepsilon(k) = x^{(0)}(k) - \hat{x}^{(0)}(k)$$

$$S_1 = \sqrt{\frac{1}{n} \sum_{k=1}^n (x^{(0)}(k) - \bar{x})^2} = 2192.751$$

$$S_2 = \sqrt{\frac{1}{n} \sum_{k=1}^n (\varepsilon(k) - \bar{\varepsilon})^2} = 543.717$$

$$\text{其中: } \bar{x} = \frac{1}{n} \sum_{k=1}^n x^{(0)}(k), \bar{\varepsilon} = \frac{1}{n} \sum_{k=1}^n \varepsilon(k), k = 1, 2, \dots$$

从而得到：

$$C = \frac{S_2}{S_1} = 0.248$$

计算小误差概率 p ：

$$p = P(|\varepsilon(k) - \bar{\varepsilon}| < 0.6745S_1) =$$

$$P(0.416, 1020.781, 927.105, 584.542,$$

$$305.204, 60.456, 28.451, 152.825 < 1479.011) = 1$$

(以上结果为方便均保留三位小数)

将计算所得 C 值和 P 值与国际通用的精度检验等级参照表对比，可得模型的精度等级为一级 ($C \leq 0.35$ 或 $P \geq 0.95$)，说明模型的检验精度很高，可用此模型进行指数预测。

8) 预测。运用前文所得时间响应序列，将所需预测事件的 k 值代入，对结果做一次累减计算，即可得未来指数预测值。此处“江歌案宣判事件”未来三小时内的热度预测值为：20040、20574、21123。

利用以上模型即可清晰看出事件发展态势及未来趋势走向，舆情相关管理部门可据此判断舆情局势，采取不同管理措施。

2.3 基于灰色关联分析的舆情事件分级过程

基于前文介绍的模型可得舆情事件未来几小时内的指数数据，但是得到这些数据尚不能完成舆情事件监管管理任务，进一步地根据所得数据对舆情事件未来发展态势进行甄别，针对事件热度等级的不同提出建议，达到提前采取行动，防止公共危机事件发生的目标。

灰色关联分析方法是本文处理舆情事件分级的主要方法。灰色关联分析是一种对系统发展变化态势进行定量描述和比较的方法，其基本思想是通过比较参考数列和若干个比较数列之间相似程度来判断其联系是否紧密。此模型所具备的样本量需求少，对无规律数据同样适用，不会出现量化结果与定性分析结果不符等情况，能够适应网络舆

情特殊环境，为判别舆情事件等级提供简便可行的方法。目前，学术界已提出了基于相似性视角和接近性视角的两种新型灰色关联分析模型，基于相似性视角的模型主要用于衡量参考序列在几何形状上与待评价数据的相似程度，基于接近性视角的模型主要用于衡量参考序列在空间中所处位置上与待评价数据的接近程度^[26]。

舆情事件分级的基本思路是：首先定义舆情事件的四个等级“蓝色”“黄色”“橙色”“红色”。其中“蓝色”代表舆情事件的发展态势不是很紧急，不需过度进行干涉或防范；“黄色”代表相应的舆情事件已有扩大的苗头，需提高警惕；“橙色”代表相应的舆情事件已非常紧急，相应的措施应尽快实施起来；“红色”则代表舆情已大范围扩散，必须采取强力措施进行舆情干涉。其次，通过经验总结及专家评估的方式给出舆情事件四个等级分别对应的基准值。最后，将一个舆情事件某个时刻的三个指数值的序列数据作为待评价数列，通过比较四个等级的基准值与待评价数列接近程度，得出舆情事件的等级。此处思想与基于接近性视角的灰色关联模型相近，因此以此模型作为舆情事件分级的方法。

此处以携程虐童案事件 2017 年 11 月 8 日 11 时的数据为例介绍分级具体过程。

1) 基准值确定。通过对以往舆情事件及相关专家意见的总结，初步得出“蓝色”“黄色”“橙色”“红色”四个等级舆情事件的基准值分别为：(6000, 5000, 400k)；(10000, 8000, 550k)；(15000, 11000, 700k)；(18000, 14000, 850k)，分别记作 $X_1 = (6000, 5000, 400)$ ； $X_2 = (10000, 8000, 550)$ ； $X_3 = (15000, 11000, 700)$ ； $X_4 = (18000, 14000, 850)$ 。

2) 待评价数据的确定。选择 2017 年 11 月 8 日 11 时携程虐童案的数据为待评价数列，记作 $X_0 = (10872, 5471, 457.8)$ 。

3) 数据无量纲化。根据数据的特点，本文使用初值化法实现数据的无量纲化过程，可得 $X_0 = (0.604, 0.391, 0.539)$ ； $X_1 = (0.333, 0.357, 0.471)$ ； $X_2 = (0.556, 0.571, 0.647)$ ； $X_3 = (0.833, 0.786, 0.823)$ ； $X_4 = (1, 1, 1)$ （为计算方便保留三位小数）。

4) 差值计算。

求 $|S_0 - S_1|$ 、 $|S_0 - S_2|$ 、 $|S_0 - S_3|$ 、 $|S_0 - S_4|$ 可得：

$$|S_0 - S_1| = \left| \frac{1}{2} [x_0(1) - x_1(1)] + \sum_{k=2}^{n-1} x_0(k) - x_1(k) + \frac{1}{2} [x_0(n) - x_1(n)] \right| = 0.203$$

$$|S_0 - S_2| = \left| \frac{1}{2} [x_0(1) - x_2(1)] + \sum_{k=2}^{n-1} x_0(k) - x_2(k) + \frac{1}{2} [x_0(n) - x_2(n)] \right|$$

$$\left| \frac{1}{2} [x_0(n) - x_2(n)] \right| = 0.211$$
$$|S_0 - S_3| = \left| \frac{1}{2} [x_0(1) - x_3(1)] + \sum_{k=2}^{n-1} x_0(k) - x_3(k) + \frac{1}{2} [x_0(n) - x_3(n)] \right| = 0.652$$
$$|S_0 - S_4| = \left| \frac{1}{2} [x_0(1) - x_4(1)] + \sum_{k=2}^{n-1} x_0(k) - x_4(k) + \frac{1}{2} [x_0(n) - x_4(n)] \right| = 1.038$$

5) 灰色关联度值计算。计算灰色接近关联度 ρ_{01} 、 ρ_{02} 、 ρ_{03} 、 ρ_{04} 。

$$\rho_{01} = \frac{1}{1 + |S_0 - S_1|} = 0.831$$
$$\rho_{02} = \frac{1}{1 + |S_0 - S_2|} = 0.826$$
$$\rho_{03} = \frac{1}{1 + |S_0 - S_3|} = 0.605$$
$$\rho_{04} = \frac{1}{1 + |S_0 - S_4|} = 0.491$$
$$\rho_{01} > \rho_{02} > \rho_{03} > \rho_{04}$$

结果表明，序列 4 也就是代表舆情等级“蓝色”的序列和待评价序列最相近，所以 2017 年 11 月 8 日 11 时

表 3 预测过程的关键数据展示

		<i>a</i> 值	<i>u</i> 值	时间响应序列	<i>C</i> 值	<i>P</i> 值
江歌案宣判	新浪微指数	-0.026332	16126.085341	$624651.90432e^{0.026332k} - 612414.90432$	0.265	1
	百度指数	-0.011887	20133.390943	$1710298.844424e^{0.011887k} - 1693729.844424$	0.386	0.875
	头条指数	0.003489	646.381025	$-184841.02912e^{-0.003489k} + 185262.32912$	0.25	1
携程虐童案	新浪微指数	-0.074064	9264.92429	$131347.699193e^{0.074064k} - 125093.699193$	0.106	1
	百度指数	-0.036818	5042.495181	$140893.288246e^{0.036818k} - 136958.288246$	0.092	1
	头条指数	-0.042315	432.376715	$10549.759067e^{0.042315k} - 10218.159067$	0.224	1
莫焕晶案宣判	新浪微指数	0.119652	15665.189607	$-120564.275363e^{-0.119652k} + 130923.275363$	0.211	1
	百度指数	0.096399	9589.937462	$-93759.212711e^{-0.096399k} + 99481.212711$	0.253	1
	头条指数	0.036685	571.47121	$-15105.538914e^{-0.036685k} + 15577.838914$	0.082	1

利用每个模型时间响应序列，即可得未来三小时内的指数预测值，结果见表 4。

3.2 分级过程

将原始数据指数值与得到的预测值代入 2.3 节所介绍的灰色关联分析法模型中，可对每个时刻的舆情事件等级进行分级。经计算，每个时刻的关联度值如表 5 所示。

利用计算所得的灰色关联度，即可判断舆情事件在某个时刻的舆情等级，舆情等级是做出舆情管理决策的重要信息，而灰色关联分析法具有很好的分类精确性，在处理分级问题上效果显著。

3.3 结果说明与管理建议

将 3.1 节与 3.2 节所得结果进行综合，即可得最终结果，如表 4 所示。

携程虐童案的舆情等级为“蓝色”。
舆情的热度等级是管理舆情的重要决策依据，政府等可据此提出不同等级下舆情应对方法，提高对舆情事件分级管理的效率。
将上述两部分介绍的方法结合，即可构建网络舆情事件预测和分级管理体系，通过此体系即可实现对网络舆情事件的科学预测和分级管理。

3 实证研究

以三个典型舆情事件“江歌案宣判”“携程虐童案”“莫焕晶案二审宣判”为例，进行数据分析（原始数据见表 2）。

3.1 预测过程

以各事件的某指数值的时间序列作为原始数据列入前文 2.2 节所介绍模型，经计算可得每个模型的关键数据，如表 3 所示。
将表 3 中所得 *C* 值与 *P* 值与国际通用的精度检验等级参照表对比，可见所有模型的检验精度水平都在二级（ $C \leq 0.5$ 或 $P \geq 0.8$ ）及以上，实证结果可以说明 EGM(1,1) 模型在指数模拟和预测方面有很好的精确度，能够有效处理舆情预测相关问题。

根据表 4 中数据可得结论如下：

1) 江歌案宣判事件。江歌案宣判事件自 14 时起就达到很高的舆论热度，舆情等级为橙色，在随后两个小时内更是迅速发展，每小时平均增速大概在 15% 左右，并于 15 时达到最高等级的红色。随后舆情热度虽未快速增长，平均增速保持在 3% 左右，但也持续保持着“红色”等级高危态势。在未来三个小时内，江歌案宣判事件的新浪微指数和百度指数虽有增长，但是增幅不大，增长幅度都不超过 3%，头条指数微有降低，但降低的幅度也不大，舆情等级依旧持续表现为红色，说明江歌案宣判事件未来几个小时内将始终保持高危态势。

江歌案宣判事件代表那些事件爆发突然，传播速度快，舆论热度大的舆情事件，事件发生期间通常会产生对

表 4 各輿情事件最终结果数据

		实际值								预测值		
时间		14 时	15 时	16 时	17 时	18 时	19 时	20 时	21 时	22 时	23 时	24 时
江歌案 宣判	三大 指数值	12237	15646	18039	18153	17732	18458	19041	19367	20040	20574	21123
		16569	20280	22032	20467	20142	21069	21982	22432	22226	22492	22760
		421. 3k	622. 7k	651. 4k	678. 8k	623. 9k	619. 7k	622. 1k	641. 1k	628k	626k	623k
	輿情等级	橙色	红色	红色	红色	红色	红色	红色	红色	红色	红色	红色
时间		9 时	10 时	11 时	12 时	13 时	14 时	15 时	16 时	17 时	18 时	19 时
携程 虐童案	三大 指数值	6254	9631	10872	12078	12935	13391	14937	15406	16957	18261	19665
		3935	5163	5471	5824	5964	6139	6319	6543	6837	7093	7359
		331. 6k	436. 2k	457. 8k	528. 7k	535. 1k	546. 7k	560. 4k	571. 9k	613k	639k	667k
	輿情等级	蓝色	蓝色	蓝色	黄色	黄色	黄色	黄色	黄色	黄色	黄色	橙色
时间		10 时	11 时	12 时	13 时	14 时	15 时	16 时	17 时	18 时	19 时	20 时
莫焕 晶案 宣判	三大 指数值	10359	12937	13038	11022	9064	8264	7258	6841	5883	5220	4631
		5722	8124	8083	7465	6953	5596	5107	4668	4387	3984	3618
		472. 3k	546. 4k	526. 7k	503. 8k	482. 3k	467. 1k	455. 3k	439. 8k	420k	405k	391k
	輿情等级	黄色	黄色	黄色	黄色	黄色	蓝色	蓝色	蓝色	蓝色	蓝色	蓝色

表 5 分级过程的关键数据展示

		实际值								预测值		
时间		14 时	15 时	16 时	17 时	18 时	19 时	20 时	21 时	22 时	23 时	24 时
江歌案 宣判	ρ_{01}	0. 497	0. 401	0. 371	0. 387	0. 394	0. 381	0. 37	0. 363	0. 363	0. 359	0. 355
	ρ_{02}	0. 626	0. 481	0. 438	0. 456	0. 47	0. 453	0. 436	0. 427	0. 427	0. 421	0. 415
	ρ_{03}	0. 864	0. 611	0. 542	0. 571	0. 594	0. 566	0. 54	0. 525	0. 526	0. 518	0. 509
	ρ_{04}	0. 813	0. 800	0. 686	0. 732	0. 77	0. 723	0. 683	0. 659	0. 661	0. 647	0. 633
	等级	橙色	红色	红色	红色	红色	红色	红色	红色	红色	红色	红色
时间		9 时	10 时	11 时	12 时	13 时	14 时	15 时	16 时	17 时	18 时	19 时
携程 虐童案	ρ_{01}	0. 901	0. 882	0. 831	0. 767	0. 746	0. 728	0. 695	0. 679	0. 641	0. 613	0. 586
	ρ_{02}	0. 657	0. 781	0. 826	0. 9	0. 932	0. 961	0. 977	0. 944	0. 872	0. 821	0. 774
	ρ_{03}	0. 509	0. 581	0. 605	0. 644	0. 66	0. 675	0. 705	0. 723	0. 773	0. 817	0. 87
	ρ_{04}	0. 426	0. 475	0. 491	0. 516	0. 526	0. 535	0. 554	0. 566	0. 595	0. 621	0. 651
	等级	蓝色	蓝色	蓝色	黄色	黄色	黄色	黄色	黄色	黄色	黄色	橙色
时间		10 时	11 时	12 时	13 时	14 时	15 时	16 时	17 时	18 时	19 时	20 时
莫焕 晶案 宣判	ρ_{01}	0. 823	0. 666	0. 671	0. 726	0. 786	0. 873	0. 93	0. 977	0. 967	0. 917	0. 876
	ρ_{02}	0. 834	0. 919	0. 929	0. 964	0. 877	0. 788	0. 747	0. 719	0. 69	0. 665	0. 643
	ρ_{03}	0. 61	0. 739	0. 733	0. 676	0. 632	0. 585	0. 562	0. 546	0. 529	0. 514	0. 501
	ρ_{04}	0. 494	0. 575	0. 571	0. 536	0. 508	0. 477	0. 462	0. 451	0. 439	0. 429	0. 42
	等级	黄色	黄色	黄色	黄色	黄色	蓝色	蓝色	蓝色	蓝色	蓝色	蓝色

社会问题激烈讨论现象，重点可能引发群众极端负面情绪泛滥，爆发群体极化危机。对此政府等輿情监管管理部门应该采取及时控制的方法，敦促监管部门加快舆论源问题处理速度，利用在群众中培植的舆论导向因子及时把控舆论方向，抑制輿情传播。

2) 携程虐童案。携程虐童案事件自 9 时起，到 11 时都是蓝色等级，但三大指数值均有大约 8% 的增长，这意味着輿情事件正处在扩散过程中，尚未造成严重后果。12 时至 16 时輿情事态已发展为黄色，但三个指数数值增长速度放缓，说明舆论事件的扩散已引起社会的关注。在未来的三小时内，三大指数值都将保持着增长态势，最大增幅在 7% 左右，在 19 点时甚至会发展为红色，说明携程事件还在持续发酵，輿情危机将会越来越严重。

携程虐童事件具有典型性，它代表那些阶段性明显，热度随事件报道逐步变化，輿情等级不断切换的輿情事件。伴随着事件深入发展，舆论多阶段爆发、剧情反转不定、造谣者众多等现象往往会频繁发生。对此，建议政府部门严厉打击造谣者，在事件早期实现对舆论的控制，对事件发展新方向保持敏感，以防輿情子话题漂移产生新的輿情危机的情况。

3) 莫焕晶案宣判。莫焕晶案宣判事件在最初发生的 5 个小时内輿情等级为黄色，但随着时间的流逝，輿情事件热度逐渐下降，下降均度在 9% 左右，并于 15 时起降低至蓝色等级，预测的三个小时内热度也都在下降，等级也一直是蓝色。所以如果不出现事件再度反弹的话，此事件热度已不足以产生公共危机事件，也不需要严加管控。

莫焕晶案宣判事件则代表那些事件持续时间不长,影响范围较小的舆情事件。对于此类事件,政府部门只要做到及时监控,敦促事件尽快处理,就能及时解决舆情控制难题。

4 结束语

本文以灰色系统理论为基础,以新浪微指数、百度指数、头条指数为指标量,构建了一套网络舆情预测与分级管理体系。通过对三大指数的应用,解决了现有体系构建成本高、数据采集困难且数据实时性差的难题,并针对网络舆情不确定性更大、传播速度更快、反应时间更短所带来的问题,运用灰色系统理论,将具有良好预测性的EGM(1.1)模型及具有良好分类准确性的灰色关联分析法相结合,实现了对舆情事件预测与分级管理两大步骤的综合运用。最后通过对三个典型舆情事件“江歌案宣判”“携程虐童案”“莫焕晶案二审宣判”的案例研究,验证了本文所提方法的合理性和准确性,并结合舆情事件的特点给出了相应管理建议。在网络舆情监管形势日益严峻的大环境下,本文提出了一套建构简便而又科学有效的网络舆情预测与分级管理体系,为提高政府的网络舆情监管效率作出了贡献,对于保护国家信息安全,构建和谐网络环境具有重要意义。

本文研究还存在一些局限:首先,三大指数具有很明显的各自公司的特点,以三大指数为指标量仅能反映舆情事件在微博、百度、今日头条的发展形势,若要全面反映舆情态势需进一步完善现有网络舆情指标体系。其次,本文虽能得到各舆情事件在不同时刻的热度及相应的舆情等级,但并未总结不同类型舆情事件的不同发展特点,也未总结针对不同舆情情况的管理方法,后续研究应进一步研究舆情事件的发展规律,建立完善的舆情应对方法体系。□

参考文献

[1] 张艳丰,李贺,彭丽徽,等.基于语义隶属度模糊推理的网络舆情监测预警实证研究[J].情报理论与实践,2017,40(9):82-89.

[2] 刘思峰,党耀国,方志耕,等.灰色系统理论及其应用[M].北京:科学出版社,2005:13.

[3] 李文杰,化存才,何伟全,等.网络舆情信息的综合评价指标体系构建与模糊评判模型[J].情报科学,2015(9):93-99.

[4] 刘健,毕强,李瑞.微博舆情信息传播效果评价指标体系构建研究——基于模糊数据包络分析法[J].情报理论与实践,2016,39(12):31-38.

[5] 王高飞,李明.基于AHP-模糊综合分析的移动社交网络舆情预警模型研究[J].现代情报,2017,37(1):41-44.

[6] 刘毅.基于三角模糊数的网络舆情预警指标体系构建[J].统计与决策,2012(2):12-15.

[7] 梁晓贺,田儒雅,吴蕾,等.基于超网络的微博舆情主题挖

掘方法[J].情报理论与实践,2017,40(10):100-105.

[8] 田世海,吕德丽.改进潜在语义分析和支持向量机算法用于突发安全事件舆情预警[J].数据分析与知识发现,2017,1(2):11-18.

[9] 郭韧,李红,陈福集.基于可拓聚类的网络舆情演化预测研究[J].情报理论与实践,2017(1):83-87.

[10] 张一文,齐佳音,方滨兴,等.基于贝叶斯网络建模的非常规危机事件网络舆情预警研究[J].图书情报工作,2012,56(2):76-81.

[11] 胡悦,王亚民.基于模糊神经网络的微博舆情趋势预测方法[J].情报科学,2017,35(12):28-33.

[12] 何建民,李雪.面向微博舆情演化分析的隐马尔科夫模型研究[J].情报科学,2016,34(4):7-12.

[13] YANG H L, LIN Q F. Opinion mining for multiple types of emotion-embedded products/services through evolutionary strategy[J]. Expert Systems with Applications, 2018, 99: 44-55.

[14] 安璐,吴林.融合主题与情感特征的突发事件微博舆情演化分析[J].图书情报工作,2017(15):120-129.

[15] 赵晓航.基于情感分析与主题分析的“后微博”时代突发事件政府信息公开研究——以新浪微博“天津爆炸”话题为例[J].图书情报工作,2016,60(20):104-111.

[16] 董坚峰.基于Web挖掘的突发事件网络舆情预警研究[J].现代情报,2014,34(2):43-47.

[17] 李彤,宋之杰.基于模型集成的突发事件舆情分析与趋势预测研究[J].系统工程理论与实践,2015,35(10):2582-2587.

[18] 王英,龚花萍.基于情感维度的大数据网络舆情情感倾向性分析研究——以“南昌大学自主保洁”微博舆情事件为例[J].情报科学,2017(4):37-42.

[19] GOPALAKRISHNAN V, RAMASWAMY C. Patient opinion mining to analyze drugs satisfaction using supervised learning[J]. Journal of Applied Research & Technology, 2017, 15(4): 311-319.

[20] CHERIF W, MADANI A, KISSI M. Towards an efficient opinion measurement in Arabic comments[J]. Procedia Computer Science, 2015, 73: 122-129.

[21] 陈涛,林杰.基于搜索引擎关注度的网络舆情时空演化比较分析——以谷歌趋势和百度指数比较为例[J].情报杂志,2013(3):7-10.

[22] 王康,李含伟.自媒体时代的企业网络舆情应对策略研究——基于上市公司百度指数的研究[J].情报科学,2018,36(1):113-117.

[23] 张玉强.基于微指数变化曲线的阶段性事件识别及演化关系分析[D].青岛:山东科技大学,2017.

[24] 刘思峰,杨英杰.灰色系统研究进展(2004—2014)[J].南京航空航天大学学报,2015,47(1):1-18.

[25] 钱吴永,党耀国.基于振荡序列的GM(1,1)模型[J].系统工程理论与实践,2009,29(3):151-156.

[26] 刘思峰,谢乃明,等.基于相似性和接近性视角的新型灰色关联分析模型[J].系统工程理论与实践,2010,30(5):881-887.

作者简介:王宁,男,1958年生,博士,教授。赵胜洋,男,1995年生,硕士。单晓红,女,1976年生,博士,副教授。

录用日期:2018-08-13